Leukemia (2013) 27, 370–376
© 2013 Macmillan Publishers Limited All rights reserved 0887-6924/13
www.nature.com/leu

ORIGINAL ARTICLE

Flow cytometry and IG/TCR quantitative PCR for minimal residual disease quantitation in acute lymphoblastic leukemia: a French multicenter prospective study on behalf of the FRALLE, EORTC and GRAALL

R Garand[1,5], K Beldjord[2,5,15], H Cavé[3,15], C Fossat[4,15], I Arnoux[4], V Asnafi[5], Y Bertrand[6], M-L Boulland[7], C Brouzes[5], E Clappier[3], E Delabesse[8], T Fest[7], F Garnache-Ottou[9], F Huguet[8], M-C Jacob[10], E Kuhlein[8], S Marty-Grès[11], A Plesa[12], N Robillard[1], M Roussel[7], J Tkaczuk[8], H Dombret[2], E Macintyre[5], N Ifrah[13], MC Béné[14] and A Baruchel[3]

Minimal residual disease (MRD) quantification is widely used for therapeutic stratification in pediatric acute lymphoblastic leukemia (ALL). A robust, reproducible, sensitivity of at least 0.01% has been achieved for IG/TCR clonal rearrangements using allele-specific quantitative PCR (IG/TCR-QPCR) within the EuroMRD consortium. Whether multiparameter flow cytometry (MFC) can reach such inter-center performance in ALL MRD monitoring remains unclear. In a multicenter study, MRD was measured prospectively on 598 follow-up bone marrow samples from 102 high-risk children and 136 adult ALL patients, using IG/TCR-QPCR and 4/5 color MFC. At diagnosis, all 238 patients (100%) had at least one suitable MRD marker with 0.01% sensitivity, including 205/238 samples (86%) by using IG/TCR-QPCR and 223/238 samples (94%) by using MFC. QPCR and MFC were evaluable in 495/598 (83%) samples. Qualitative results ($<0.01\%$ or $\geqslant 0.01\%$) concurred in 96% of samples and overall positivity (including $<0.01\%$ and nonquantifiable positivity) was concurrent in 84%. MRD values $\geqslant 0.01\%$ correlated highly ($r^2 = 0.87$) and 69% clustered within half-a-$\log_{10}$. QPCR and MFC can therefore be comparable if properly standardized, and are highly complementary. MFC strategies will benefit from a concerted approach, as does molecular MRD monitoring, and will contribute significantly to the achievement of 100% MRD informativity in adult and pediatric ALL.

Leukemia (2013) **27**, 370–376; doi:10.1038/leu.2012.234

Keywords: acute lymphoblastic leukemia; minimal residual disease; flow cytometry; real-time quantitative PCR; IG/TCR rearrangement

INTRODUCTION

Minimal residual disease (MRD) monitoring is a strong prognostic factor in acute lymphoblastic leukemia (ALL), incorporated as a powerful stratification criterion in ongoing pediatric therapeutic trials.[1–6] Recent results suggest that MRD evaluation also helps to recognize patients who may benefit from hematopoietic stem-cell transplantation in adult ALL.[7,8] Quantification of IG/TCR clonal rearrangements using allele-specific quantitative PCR (IG/TCR-QPCR)[9–13] and multiparameter flow cytometry (MFC)[4,14–16] are the most commonly used methods for MRD assessment. The EuroMRD Consortium guidelines and quality control (QC) runs have established a high degree of standardization and inter-laboratory reproducibility of MRD results obtained with molecular quantification of, preferably, two IG/TCR targets,[17–19] centralized in a limited number of experienced laboratories. Consequently, IG/TCR-QPCR assays are currently considered to be the gold standard for MRD monitoring in most European ALL therapeutic trials.[20]

MFC, based on the detection of leukemia-associated immunophenotypic (LAIP) abnormalities at diagnosis, is a more recently introduced alternative MRD technique in ALL, the specificity, sensitivity and feasibility of which has increased with the development of multiparameter analysis. By relying on stored diagnostic and follow-up information (flow-cytometry list modes) and providing complementary data, it bypasses some of the limitations of PCR-MRD measurement in terms of diagnostic DNA availability, as well as superior informativity in certain ALL subsets of immature ALL.[21] Although increasing efforts have been developed for standardization throughout Europe, inter-laboratory reproducibility of MFC-MRD results need to be assessed before this methodology can be implemented into multicenter ALL treatment schedules.[22–24] Whether MFC can compete with IG/TCR-QPCR performance in routine ALL follow-up monitoring is open-to-question, in both pediatric and adult ALL.

[1]CHU Nantes, Nantes, France; [2]APHP-Hôpital Saint Louis, Paris, France; [3]APHP-Hôpital Robert-Debré, Paris, France; [4]Hôpital de la Timone, Marseille, France; [5]Department of Hematology, APHP-Hôpital Necker, Paris, France; [6]IHOP, Lyon, France; [7]CHU Rennes, Rennes, France; [8]CHU Toulouse, Toulouse, France; [9]CHU Besançon, Besançon, France; [10]CHU Grenoble, Grenoble, France; [11]CHU Montpellier, Montpellier, France; [12]Hospices Civils de Lyon, Hôpital Edouard Herriot, Lyon, France; [13]CHU Angers, Angers, France and [14]Immunology Laboratory, Université de Lorraine & CHU Nancy, Nancy, France. Correspondence: Professor E Macintyre, Department of Hematology, Assistance Publique-Hôpitaux de Paris (APHP), Hôpital Necker-Enfants-Malades, 149 Rue de Sèvres, 75015 Paris, France.
E-mail: elizabeth.macintyre@nck.aphp.fr
or Professor M-C Béné, Immunology Laboratory, Université de Lorraine & CHU Nancy, 54500 Vandoeuvre les Nancy, France.
E-mail: bene@univ-lorraine.fr
[15]These authors contributed equally to this work.
Received 23 April 2012; revised 16 July 2012; accepted 26 July 2012; accepted article preview online 16 August 2012; advance online publication, 16 October 2012

We (and others) have previously shown that results obtained using MFC in a single-center setting were comparable to those obtained using IG/TCR-based QPCR in pediatric ALL.[21,25–27]

The objectives of the present study were to evaluate whether comparable results can also be obtained in a multicenter setting in childhood and in adult ALL.

We therefore undertook a multicenter national MRD study comparing 4–5 color MFC- and molecular-based methods in childhood and adult BCR-ABL negative ALL (Soutien aux Techniques Innovantes et Coûteuses (STIC) 2006 program). To enrich for positive MRD samples in children, only high-risk cases were included. We show that highly systematized MFC analysis and reporting is at least as informative as, and has a high degree of concordance with, the gold standard IG/TCR-QPCR technique at the targeted 0.01% threshold, in children as well as in adults. Importantly, QC results confirmed good reproducibility between MFC laboratories working with different instruments but in close collaboration.

PATIENTS AND METHODS

Patients

Between July 2006 and December 2009, 238 patients (102 children and 136 adults) from 18 French clinical centers were included in the study. Adult patients were treated according to the GRAALL-05 clinical trial. Pediatric patients were treated according to either the FRALLE-2000 or EORTC-58951 trials, and were included, provided they displayed high-risk features according to NCI criteria (age $\geqslant$10 years old; WBC count $\geqslant 50.10^9$/l) or very high-risk criteria defined by the trial (poor response to pre-phase steroid treatment qualified by the presence of $\geqslant 1.10^9$/l peripheral blasts at day 8). Patients displaying BCR-ABL-positive ALL or aged $<$1 year were excluded from the study, as these patients are treated on separate, multinational protocols in which the relative place of MCF and molecular monitoring merits specific evaluation.

The diagnosis and immunological classification of ALL were established according to EGIL criteria.[28] A B-cell precursor (BCP)-ALL immunophenotype was found in 153 patients, and a T-ALL immunophenotype in 85.

MRD studies were performed in 598 follow-up samples from these 238 ALL cases. Follow-up bone marrow samples were obtained at the end of induction (MRD1, day 35–45), after the first consolidation (MRD2, day 70–95) and at later time points in 96, 57 and 79 children and 126, 118 and 122 adults, respectively. All samples fulfilled the remission criteria ($<$5% blasts at cytomorphological examination).

Sample preparation

For each patient, malignant cells were obtained at diagnosis from bone marrow, or, in some cases with tumoral mass, from pleural fluid or lymph node aspirates.

Bone marrow samples underwent Ficoll centrifugation followed by blast percentage evaluation on cytospins. Only samples with $>$25% of blasts at diagnosis were included in the molecular study. Whenever necessary, cell pellets were stored in liquid nitrogen until use. Total RNA and genomic DNA were extracted according to standard procedures. LAIP features were studied on a fresh blood or bone marrow sample taken at diagnosis. For MRD studies, both techniques were performed in parallel on a unique bone marrow sample that was subdivided into two fractions, one immediately processed by MFC, the other one cryopreserved or extracted immediately for molecular analysis.

IG/TCR-QPCR

Clone-specific IG/TCR-QPCR was assessed and used for MRD monitoring, on the basis of the quantitative detection of leukemic clone-specific T-cell receptor or immunoglobulin gene rearrangements, in five laboratories, according to EuroMRD guidelines, as described.[16–18] MRD was thus assessed between 500 ng (from one laboratory) and 675 ng (100 000 cells) (from four laboratories) of DNA, in triplicate. Positive MRD results were quantified, if within the quantitative range, and considered as 'positive non-quantifiable', if below this range.

Multiparameter flow cytometry

Four-color (39%) or five-color (61%) MFC was performed on fresh cells (after Ficoll-separation (80%) or erythrocyte-lysis (20%)) with similar antibody panels, antibody clones and gating strategies, in eight laboratories. The BCP-ALL panel consisted of antibody combinations with a common ALB1/CD10FITC, J4-119/CD19APC, J33 or 2D1/CD45PerCP (or ECD) backbone and variable fourth (PE conjugated) and fifth (PE-cyanin-7 conjugated) antibodies against 39C1.5/CD2, L138/CD13, 80H5/CD15, B9E9/CD20, S-HCL1/CD22, ALB9/CD24, P67.6/CD33, 581/CD34, HB7/CD38, AICD58/CD58, 104D2/CD117 and 9F5/CD123 antigens. These panels were modified by adding CD20FITC to the backbone in cases with a $CD10^{neg}$ $CD20^{neg}$pro-B/B-I ALL immunophenotype. Four- and five-color panels were designed at the beginning of the study and the combinations used for MRD assessment were established and maintained throughout.

The T-ALL panel consisted of antibody combinations with a common UCHT1/surface CD3 (PE-Cyanin-7 or ECD conjugated) and UCHT1/cytoplasmic CD3 (APC or Cyanin-5 conjugated) backbone and variable third (FITC conjugated), fourth (PE conjugated) and fifth (PE-cyanin 7 conjugated) antibodies against BL6/CD1a, 39C1.5/CD2, BL1a/CD5, 8H8.1/CD7, ALB1/CD10, 581/CD34, J33 or 2D1/MY31/CD56, Tü12/CD99 and HT6/TdT antigens. Cytoplasmic CD3 and nuclear TdT labeling were performed with the IntraStain fixative and permeabilization solution (Dako, Glostrup, Denmark).

At least three or two antibody combinations were used in 200/232 (86%) and 261/366 (71%) cases for MFC-MRD studies, in 4 and 5 colors, respectively. All antibodies were purchased from Becton-Dickinson (San Jose, CA, USA), Beckman-Coulter (Miami, FL, USA) or Dako. Four-color MFC analyses were performed on FACScalibur equipped with CellQuest Pro and Paint-a-gate software (Becton-Dickinson) and 5-color MFC analyses were performed on FC500 with CXP software (Beckman-Coulter) or FACScanto-II with DIVA software (Becton-Dickinson).

According to the instrument's performance, we used either a live-gate two-step acquisition procedure with FACSCalibur (Becton-Dickinson), including a selective secondary acquisition of 1000–100 000 total viable B- or T-cells that is, a median of 440 000 total nucleated cells, as reported previously.[21] A one-step initial acquisition of at least 200 000 events (median 389 000 total viable nucleated cells) was used with FC500 (Beckman Coulter) and FACSCanto-II (BD Biosciences, San Jose, CA, USA) analyzers. This ensured achievement of the sensitivity threshold of 0.01% decided upon at the initiation of the study. Doublets/dead cells and debris were excluded on FSC/SSC and CD45/SSC dot plots. Then, a CD19$+$ or cCD3$+$ cell gating allowed selection of all lymphoid B- or T-cells (that is, normal and leukemic). Within these cells, positive MRD was identified as a significant cell cluster (that is, $>$10 events) with such LAIP and scatter properties as defined at diagnosis in all plots of interest. In case of discrepant results between antibody combinations, the final MRD value considered was the highest among combinations that most reliably discriminated abnormal blasts from normal bone marrow cells.

In this respect (see Supplementary Table 1), the most frequent antibody combinations used in this study for MRD detection in addition to the CD45, CD19, CD10 backbone included CD38, CD58, CD123 and CD34 in B-ALL cases with frequencies between the eight MFC centers of 53–79% (mean: 61%), 24–79% (mean: 57%), 35–44% (mean: 38%) and 30–44% (mean: 34%), respectively. The variation of these frequencies between laboratories was not statistically significant, except for combinations including CD58. The best combinations to measure MRD in T-ALL were those containing TdT, CD99, CD1a and CD34 with usage frequencies of 73–89% (mean: 80%), 56–81% (mean: 63%), 36–50% (mean: 37%) and 15–46% (mean: 26%), respectively.

Statistical analyses

Applicability and concordance of MRD methods were compared with a χ^2 test and the Fischer exact test for binary variables and a Mann–Whitney Rank Sum test for continuous variables, using the SigmaStat 3.5 software (Systat Software, Chicago, IL, USA). Correlations between MRD values generated by QPCR and MFC were measured with the Spearman's rank correlation test and their representation was plotted using the GraphPad Software (GraphPad Software Inc., San Diego, CA, USA). The significance limit for *P*-values was set to $P<0.05$ in all tests.

RESULTS

MFC QCs and consensus interpretation

Mock MRD samples consisting of fresh leukemic cells from six BCP-ALL and one T-ALL admixed into normal regenerative bone marrows were distributed via express mail to all MFC labs. In six of these QC samples, expected positive MRD results ranged from

© 2013 Macmillan Publishers Limited

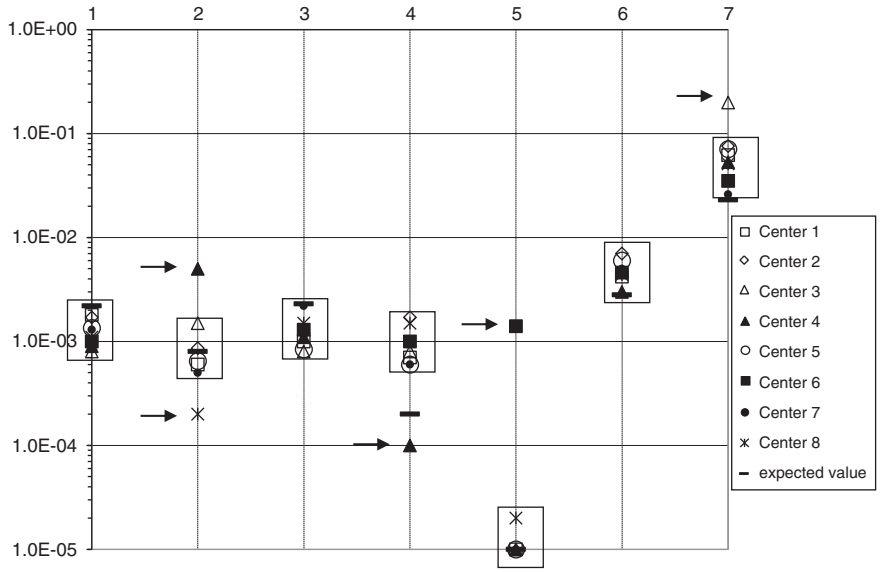

Figure 1. Flow cytometry QCs. Mock MRD samples consisted of leukemic cells spiked into fresh normal bone marrow cells. Blast cells were derived from acute lymphoblastic leukemia with CD10+ (QC number 1, 3, 5, 6 and 7) or CD10- (QC number 2) B-cell or immature T-cell (QC number 4) immunophenotype. Positive results were considered as concordant when values were clustered within a half-a-$\log_{10}$ range, which is shown by a vertical rectangle in each column. Arrows indicate off-limit results. Quality-control number 5 consisted of a normal regenerative bone marrow sample with no addition of leukemic cells. The positive value $>0.01\%$ provided by one of six laboratories was therefore a false-positive result (arrows).

0.02 to 2.30%. All of the results provided by the MFC labs were positive, that is, $\geqslant 0.01\%$, and 42/46 (91%) were clustered within a half-a-log in each QC. The lowest level of MRD tested corresponded to the only T-ALL (sample 4). It is noteworthy that all but one laboratory quantified this result at $\sim 0.1\%$, rather than the expected 0.02%, suggesting that the expected result was inaccurate. False positivity was observed in 1/6 measurements of the QC5, which was devoid of leukemic cells (Figure 1). This discrepant result may be explained by the unusually high proportion of normal immature precursor B cells present in the normal bone marrow distributed, the CD10+^{high} CD19+ CD34+^{high} CD38+$^{high/medium}$ immunophenotype of which strongly overlapped with that of the BCP-ALL cells chosen as potential contaminator in QC5.

Collective review of MFC-MRD data files was organized during five consecutive workshops in the presence of all eight laboratories in order to improve standardization of the gating strategies during the training phase of the study. These files included all those where the initial locally-issued MFC result was discordant with that of PCR-MRD ($n = 61/598$, 34 with qualitative discordance, that is, negative ($<0.01\%$) versus positive ($\geqslant 0.01\%$) results or vice versa, and 27 positive MRD results with a quantification difference higher than half-a-$\log_{10}$). In addition, 80 randomly chosen files with initially concordant MFC and PCR-MRD results were also reviewed.

MFC datafiles were collectively reanalysed with CellQuestPro, Diva (Becton-Dickinson) and CXP (Beckman-Coulter) softwares without knowledge of either the local MFC or molecular MRD results at the time of panel review. Comparison between the local and collective results is shown in Figure 2. The consensus conclusion ($<$ or $\geqslant 0.01\%$) differed from that of the local center in 19/141 (13%) cases and was retained for further analyses. As expected, differences were more frequent for cases showing an initial difference with the PCR result (17/61, 28%) than for those chosen randomly (2/80, 3%) (Table 1). The level of positivity was very low in most of these discrepancies, ranging from 0.01 to 0.03% in 17/19. All IG/TCR PCR results that initially showed a difference with MFC were reviewed and confirmed by two of the

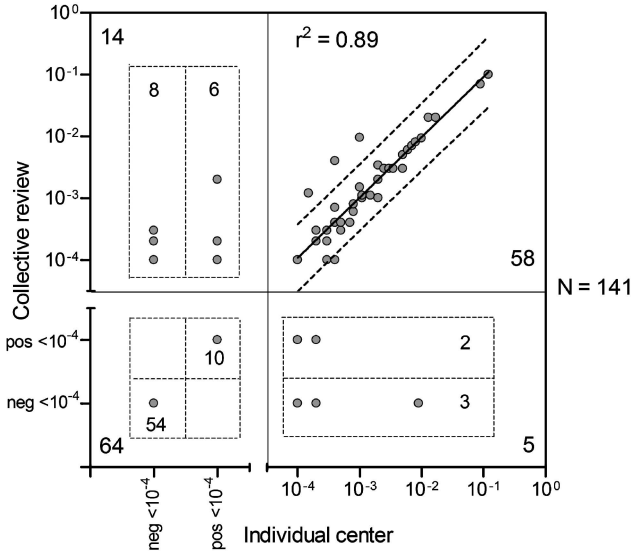

Figure 2. Concordance of MFC-MRD results provided by each individual center and after collective review of data files. The correlation coefficient of the 58 concordantly positive $\geqslant 0.01\%$ values was calculated with the Spearman's rank correlation test. All results with identical XY co-ordinates are represented as a unique dot on the figure, and numbers indicate the partition of samples in each quadrant.

molecular laboratories. The reasons for MFC-MRD discordance between local and collective results included an incorrect gating strategy ($n = 11$), poorly discriminant LAIP ($n = 6$), an unrecognized subpopulation of leukemic cells with a distinct LAIP ($n = 1$) or incomplete antibody panel review after complementary immunophenotype performance ($n = 1$). There was no center effect in these discrepancies. Comparison of 174 paired MFC-MRD analyses on Ficoll-separated mononuclear cells and erythrocyte-lysis

© 2013 Macmillan Publishers Limited

Table 1. Collective MFC-MRD file review (five meetings; review with CXP and Diva softwares)

Result of the collective review of MFC-MRD	*Local MFC-MRD result discordant with PCR-MRD*		*Local MFC-MRD result concordant with PCR-MRD*[a]		*All reviews*	
Agreement	40	66%	78	97%	118	84%
Qualitative discordance[b]	17	28%	2	3%	19	13%
Quantitative discordance[c]	4	6%	0	0%	4	3%
Total	61	43%	80	57%	141	

Abbreviations: MFC, multiparameter flow cytometry; MRD, minimal residual disease; PCR, polymerase chain reaction. [a]MFC-MRD file randomly chosen. [b]Qualitative: positive/negative (i.e. $<0.01\%/\geqslant 0.01\%$). [c]Quantitative: positivity $>1/2$ $\log_{10}$.

whole nucleated cells (from the same samples) provided 96% concordance ($<$ or $\geqslant 0.01\%$) and a Spearman's correlation factor $r^2 = 0.89$ ($P < 0.001$). On this basis, the choice of sample preparation was left to local laboratories as results were considered interchangeable, but Ficoll separation was nonetheless performed for DNA extraction by PCR.

Applicability of IG/TCR-QPCR and MFC MRD methods

IG/TCR and LAIP markers were considered suitable when they allowed MRD detection at a level of at least 0.01%. A QPCR and/or MFC target with 0.01% sensitivity was achieved in 87% and 94% of 102 children and in 86% and 96% of 136 adults, respectively. Both techniques yielded 0.01% sensitivity in 86% of children and 85% of adults and, importantly, at least one technique gave sufficient sensitivity in all cases. Overall, at least one suitable IG/TCR marker was identified at diagnosis in 205/238 (86%) patients and at least one suitable LAIP in 223/238 (94%) patients. Five patients (1 child and 4 adults) with 4–10% blasts in the diagnostic sample were excluded from molecular analysis, yet could be processed by MFC. Two or more suitable IG/TCR markers were used in 59% of pediatric and 42% of adult ALLs. Suitable IG/TCR probes were less frequently found in CD10 $-$ (13/20, 65%) than in CD10 $+$ (120/132, 91%) BCP-ALL subtypes, whether the MLL gene was rearranged (4/7, 57%, $P = 0.02$) or not (9/13, 69%, $P = 0.04$, χ^2 test). Similarly, there was significantly lower feasibility of QPCR MRD determination in immature T-I/T-II (17/26, 65%) compared with more mature T-III/T-IV (49/56, 88%) T-ALL subtypes ($P < 0.02$). By contrast, the applicability of MFC was uniformly high in CD10 $-$ (20/20, 100%) and CD10 $+$ (122/132, 92%) BCP-ALL, as well as in T-I/-TII (26/27, 96%) and T-III/T-IV (47/48, 98%) T-ALL subtypes.

Paired MRD measurements performed by both QPCR and MFC were evaluable in 495/598 (83%) bone marrow aspirates, including 196/232 (84%) and 299/366 (82%) samples from children and adults, respectively. Interestingly one or the other method could be performed in all of the remaining 103 samples, illustrating their highly complementary applicability.

Concordance of IG/TCR-QPCR and MFC-MRD values

Concordance was assessed both before and after consensus agreement on MFC results. Before consensus on MFC values and using a 0.01% threshold to define positivity, the MRD qualitative conclusion (that is, MRD negative, $<0.01\%$ or MRD positive, $\geqslant 0.01\%$) provided by both methods was identical in 461/495 samples (94%) from 203 patients The proportion of discordant files decreased from 25/236 (10.6%) for the first half of the program to 9/259 (3.5%) for the second half ($P < 0.01$), in keeping with progressive harmonization of MFC processing and interpretation.

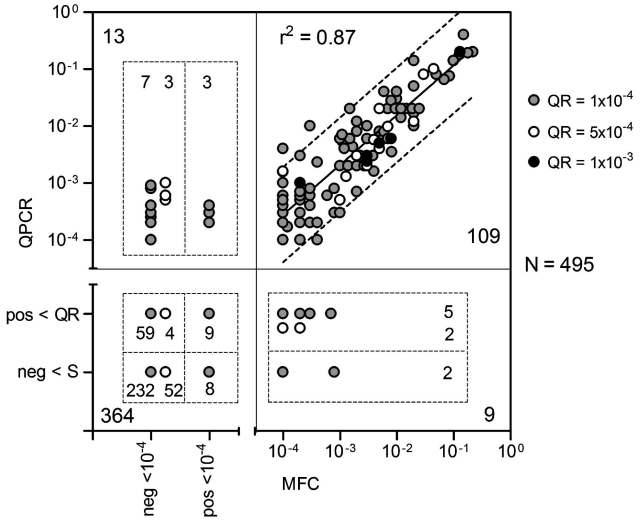

Figure 3. Concordance of 495 paired MRD results obtained by IG/TCR-RQ-PCR and MFC with a maximum sensitivity level of $\leqslant 0.01\%$ by both methods. The follow-up samples were obtained in children and adult patients after induction therapy ($n = 80$ and $n = 107$, respectively), after the first consolidation therapy ($n = 48$ and $n = 98$, respectively) or at later stages ($n = 63$ and $n = 93$, respectively). Detectable MRD below the QR is termed 'pos $<$ QR' and included six patients with a QR of 0.05% or 5×10^{-4} (unfilled circles) and 73 with a QR of 0.01% or 10^{-4}. Negative PCR results are identified as 'neg $<$ Sensitivity (S)' all had QPCR targets with a sensitivity of at least 0.01% and include 52 samples with a quantitative range of 0.05%. MRD results by QPCR with a quantitative range of 0.01% ($n = 174$), 0.05% ($n = 15$) and 0.1% ($n = 6$) are shown as gray, white and black dots, respectively. The correlation coefficient of the 109 MRD values positive $\geqslant 0.01\%$ measured by both methods was calculated with the Spearman's rank correlation test. All results with identical XY co-ordinates are represented as a unique dot on the figure, and numbers indicate the partition of samples in each quadrant.

After consensus MFC evaluation, the number of discordant results decreased from 34 to 22 and the incidence of identical MRD qualitative conclusions rose to 473/495 samples (96%) (Figure 3). Among the remaining 22 discordant results, 13 were positive with QPCR but negative with MFC and 9 vice versa. The positivity level was, however, below the molecular quantitative range of at least 0.05% in 20/22 (91%) of these samples. Further comparison was undertaken with consensus, corrected MFC results.

For molecular results, the quantifiable range (QR) refers to robust, reproducible, positivity, below which the lowlevel positivity cannot be reliably quantified. As these 'suspect' samples may include leukemic cells, they are classified as positive nonquantifiable, below QR.[19] Among the 205 cases suitable for QPCR, 176 (86%), 27 (13%) and 2 (1%) had QR of 0.01%, 0.05% and 0.1%, respectively. At the individual-sample level, of 495 samples suitable for QPCR, 416 (84%), 73 (15%) and 6 (1%) had QR of 0.01%, 0.05% and 0.1%, respectively. Overall, QR of 0.05% or 0.1% were less frequently observed in pediatric (21/195, 11%) than in adult samples (58/300, 19%, $P = 0.02$). When considering all positive MRD samples (both above and below QR for QPCR), the two methods yielded concordant MRD results in 414/495 (84%) samples. Thus, although most often (65/75) nonquantifiable, MRD was detectable with QPCR but not with MFC in 73 (14%) samples and vice versa in only 12 (2%) additional samples. It is noteworthy that six samples were positive nonquantifiable $<0.05\%$ by QPCR,

© 2013 Macmillan Publishers Limited

Table 2. Concordance of Ig/TCR-QPCR and MFC-MRD results according to age and time point

Paired MFC and PCR-MRD age and time points	*Feasible*	*Qualitative concordance*	*Paired positive RQ-PCR and MFC-MRD*	*PCR+/MFC+ MRD ratio ⩽3*
Children, all time points	191	180 (94%)	35 ($r^2=0.85$)[a]	25 (71%)
Children, post-induction	80	76 (95%)	18 ($r^2=0.84$)[a]	14 (78%)
Children, post-consolidation 1	48	45 (94%)	8	4 (50%)
Adults, all time points	298	289 (97%)	74 ($r^2=0.87$)[a]	50 (68%)
Adults, post-induction	107	102 (95%)	38 ($r^2=0.81$)[a]	24 (63%)
Adults, post-consolidation 1	98	96 (98%)	16 ($r^2=0.86$)[a]	13 (81%)

Qualitative concordance: MRD concordantly negative (<0.01%) or positive (⩾0.01%) with PCR and MFC methods. [a]Spearman's rank correlation coefficient of MRD quantification with PCR and MFC.

and as such could not be stratified at the 0.01% cutoff used in most clinical trials. Of these, two were positive by CMF at 0.01% and 0.03% and four gave CMF values <0.01%. As such, the CMF/QPCR complementarity allowed clarification of the small minority of samples with borderline values owing to insufficiently sensitive QR.

Quantification of the 109 paired MRD results positive with both QPCR and MFC were significantly correlated ($r^2=0.87$, $P<0.0001$) and clustered within half-a-$\log_{10}$ in 75 (69%) of these. By comparison, a similar proportion of paired positive results before consensus correction were within half-a-$\log_{10}$ (67/101, 67%).

The qualitative concordance of consensus paired QPCR- and MFC-MRD results was comparable in children (180/191 (94%)) and in adults (289/298 (97%)) (Table 2). Quantified positive MRD values with both techniques were strongly correlated in children ($n=35$, $r^2=0.85$, $P<0.001$) and in adults ($n=74$, $r^2=0.87$, $P<0.0001$). In addition, similar proportions of paired MRD results were clustered within half-a-$\log_{10}$ in both age groups (25/35 (71%) and 50/74 (68%), respectively).

According to follow-up time points (Table 1), the qualitative concordance of both MRD1 and MRD2 results provided by both methods was not different in children (76/80 (95%) and 45/48 (94%), respectively), nor in adults (102/107 (95%) and 96/98 (98%), respectively).

QPCR and MFC results were compared using a QPCR/MFC-MRD ratio. The percentage of paired results that were within half-a-$\log_{10}$ at both time points was not statistically different at MRD1 (24/38 (63%) nor at MRD2 (13/16 (81%)).

The proportion of identical qualitative QPCR and MFC-MRD conclusion was similar in BCP- (304/318 (96%) and in T- (165/171 (96%)) ALL subtypes, as well as with 4-color (181/189 (96%)) or 5-color (288/300 (96%)) MFC. Quantitative agreement of paired positive MRD results with both techniques, as reflected by the percentage of those clustered within half-a-$\log_{10}$, was also comparable in BCP- (57/80 (71%)) and T- (18/29 (62%)) ALL, as well as with 4- (25/35 (71%)) or 5-color (50/74 (68%)) MFC analyzers. The incidence of qualitative (that is, negative <0.01% versus positive ⩾0.01%), as well as quantitative (that is, positivity ratio >1/2 $\log_{10}$) discordant results provided by both techniques were not significantly different when molecular quantification of MRD was performed with one instead of two or more IG/TCR targets (9/241, 3.9% versus 11/240, 4.6% and 11/49, 22% versus 21/57, 37%, respectively).

Although the number of samples studied by each of the participating MFC and QPCR laboratories was variable, and some only studied children (1 MFC laboratory and 1 QPCR laboratory) while others studied only adults (two MFC laboratories and one QPCR laboratories, see Supplementary Table 2), the qualitative MRD concordance between QPCR and MFC was not different between MFC laboratories (94–100%) nor between QPCR laboratories (95–98%). Of note, discordances between both MRD techniques were not clustered as they were observed in all participating MFC and PCR laboratories before, and in 6/8 MFC and 5/5 PCR laboratories after consensus review of the files.

DISCUSSION

QPCR is currently recognized as the gold standard for MRD detection in ALL, yet increasing reports indicate that MFC also yields relevant information. Here we compared these two techniques in parallel on partitioned bone marrow samples from 238 ALL patients, and confirmed the good concordance of these methods. The decision to analyze only high-risk pediatric patients may have introduced a bias in terms of frequency of positive MRD. This was deliberate, in order to obtain enough positive samples for proper comparisons between the methods tested.

To be incorporated into therapeutic trials, standardization and QC are the required conditions for all MRD techniques. Accordingly, the EuroMRD consortium has developed highly powerful guidelines and quality assurance that guarantee reliability and reproducibility of clone-specific IG-TCR-QPCR results and makes it the gold standard method for MRD measurement.[18–20] However, several groups have shown that standardization of MFC-MRD evaluation is feasible in a multi-center setting.[22–24] In the present study, standardization included similar antibody panels and clones, with fixed backbone triplets in 4- and 5-color combinations. The choice of fluorochromes had, however, to be adapted to the constraints induced by each of the three flow cytometers in usage in the MFC laboratories network at the start of the program. Agreement on gating strategies for MRD determination also remained a constant objective during the development of this program and was improved by biannual workshops organized to that aim. During these workshops, 141 MFC-MRD data files from 81 patients were thus collectively reviewed and the consensus concordance between the initial local center and collective reanalyzed results appeared high (87%), especially for randomly chosen files. Using comparable inter-laboratory tests of list-mode data interpretation, Dworzak *et al.*[22] reported a very high degree of agreement among four centers despite differences in instruments and software. With a similar approach, Bjorklund *et al.*[24] also obtained a high concordance of MRD results between 15 laboratories at a cut-off level, which will be applied for clinical decisions, with continuing standardization resulting in better concordance.

Between-centers reproducibility was also tested by providing QC mock-MRD samples prepared by spiking ALL blasts into regenerative normal bone marrow cells. Irving *et al.*[23] had used a similar approach by testing 15 mock-MRD exchanged fresh samples, and we also obtained a high inter-laboratory agreement with 91% of the 46 positive values clustered within half-a-$\log_{10}$ in each QC. Another study also reported a high concordance (ICC 0.98) between 4 laboratories from Italy, Germany and Austria by testing artificial serial dilution preparations of fresh ALL cells admixed into normal bone marrows.[22] In our study, however, a false-positive result was noted, due to the strong immunophenotypic resemblance between leukemic cells and the immature normal B-cell precursors present in an unexpected high proportion in this QC sample. This low incidence of false-positive/negative qualitative results is thus comparable to that by Dworzak *et al.*[22] who

© 2013 Macmillan Publishers Limited

observed five similar qualitative discordances among 164 MRD values available in their study.

The high applicability levels of both IG/TCR-QPCR (86%) and MFC (94%) MRD techniques observed in our study are comparable to the most recently reported series.[4–7,25,27,29] Accordingly, all MRD determinations were feasible with at least one of the methods. This was of particular interest in certain ALL subsets, including B-I CD10- ALL or immature T-cell ALL in which MFC techniques can bypass the limitations of PCR.[21] Studies of other genes (TAL deletions, IGK, TCRβ), which were not performed systematically in the present study, could also increase the feasibility of PCR.[11,12,17] Recently, Coustan-Smith *et al*.[30] have also shown that the incorporation of new markers in antibody combinations could substantially improve the applicability and sensitivity of MRD MFC studies in BCP-ALL. Thus all these methods appear complementary, each with specific strengths and potential weaknesses.[29]

In line with most of the published reports comparing QPCR and MFC performance for MRD assessment in ALL, we chose a cut-off of 0.01% to define MRD positivity.[23,25–27,31,32] The overall qualitative concordance rate, that is, similar positive or negative MRD conclusion with both techniques, was remarkably high (96%) in our study and not different in children (94%) or adults (97%), nor according to follow-up time points. Results previously reported in the literature in childhood series ranged from 75 to 97% concordance, some with similar rates, whatever the time point,[25,26] others with a lower rate at early time points.[23,32] We report comparable results in adult samples. Several differences between the series may explain these discrepancies. Multi-centric experiences reported lower concordance rates (75–86%)[23,31,32] than mono-centric studies (89–97%),[25–27] although significant standardization efforts for MRD processing were developed in the former. Our study was multicentric, involving, in particular, a relatively high number of participating MFC laboratories. Harmonization of antibody panels and gating strategies together with collective MFC data file review improved our results, as 28% of the discrepancies between QPCR and MFC MRD qualitative conclusions disappeared after collective data reanalysis.

Usage of a unique bone marrow sample to perform each of paired QPCR and MFC-MRD analysis may also have contributed to the higher qualitative agreement of our results, similarly to what Neale *et al*.[25] and Kerst *et al*.[27] reported, by comparison to the three published series that used distinct samples to perform dual MRD analysis.[23,31,32]

The presence of qualitative discrepant MRD results in our study deserves several observations: First, similarly to previous reports,[23,25,31] the positivity level was most often very low and around the threshold level, second, most discordant paired samples positive in MFC contained detectable nonquantifiable (mostly $<0.01\%$) MRD signals with QPCR. In a minority of these, with a molecular quantitative range of only 0.05%, the CMF result allowed classification above or below the 0.01% level used for clinical stratification. On the other hand, MFC failed to detect any MRD signal in the majority of discordant QPCR-MRD positive samples. The former type of discrepancy is likely to reflect difficulties encountered with QPCR to accurately and reproducibly quantify the lowest level of MRD according to Poisson's distribution, whereas the latter is more likely to be either due to the lower sensitivity of MFC methods or due to the fact that it analyses only live cells. This is further supported by our observation that 12% of paired MRD results classified as negative with both methods indeed contained detectable leukemic signals at a level below the threshold with QPCR but not with MFC.

However, the clinical significance of detectable abnormal cells at a level $<0.01\%$ is still questionable. First of all, non-specific PCR amplification of normal DNA may be confused with low-level MRD signals. In addition, various proportions of ALL cells detected by QPCR had undergone apoptosis and were therefore excluded from MFC analysis. Similarly, almost all leukemia-associated markers can be expressed by small subsets of normal lymphoid precursor cells, which could lead to misinterpretation of very low MFC-MRD signals. A recent study suggested that the detection of a very low level of MRD (0.001–0.01%) with QPCR methods at the end of remission induction therapy has a prognostic significance in childhood ALL,[33] but other studies are requested for confirmation. Nevertheless, given its higher capacity to detect low MRD levels, QPCR may be preferable to MFC to prevent false-negative MRD results in trials aiming to recognize low-risk patients in which treatment de-escalation may be intended. As most of the discordant undetectable MFC-MRD results were seen in samples qualified positive by QPCR, but at a level around 0.01%, MFC would be an appropriate MRD method to identify high-risk patients in whom treatment intensification may be necessary, given the rapidity with which the results can be obtained in routine practice. In particular, MFC may be especially valuable to evaluate the chemo-sensitivity of the leukemic clone by early MRD assessment during remission induction therapy.[34–36]

Similarly to all previous reports,[23,25–27,31] the quantification levels of our 109 MRD samples positive with both QPCR and MFC were correlated.

The reliability of MRD values may be lowered by other parameters, which are highly difficult to detect and quantify in routine practice. They include the degree of contamination of bone marrow samples with peripheral blood, presence of minor immunophenotypic subpopulations with distinct LAIP, minor molecular subclones, unexpected changes of clonal markers and antigenic modulation during treatment or at early relapse.[20,25,31,37–40] Therefore, MRD values obtained with these two techniques are not yet easily exchangeable.[26] As a consequence, positivity thresholds to define risk categories have to be established in each treatment protocol according to the methods chosen to assess MRD monitoring and their aims.

In conclusion, this multicentre study confirms that QPCR and MFC are highly complementary strategies for MRD detection in ALL, likely to provide 100% informativity in both adult and pediatric ALL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by a grant 'Soutien aux Techniques Innovantes et Coûteuses' (STIC) by the French Institut National du Cancer (InCA) and benefited from support by the companies BD Biosciences and Beckman Coulter. We are grateful to all clinicians, biologists and clinical research assistants involved in the FRALLE, EORTC and GRAALL clinical trials, where the patients were enrolled. Special thanks are addressed to Véronique Lhéritier for her great help with the adult patients'database.

REFERENCES

1 van Dongen JJ, Seriu T, Panzer-Grumayer ER, Biondi A, Pongers-Willemse MJ, Corral L *et al*. Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. *Lancet* 1998; **352**: 1731–1738.
2 Cavé H, van der Werff ten Bosch J, Suciu S, Guidal C, Waterkeyn C, Otten J *et al*. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia. European organization for research and treatment of cancer – childhood leukemia cooperative group. *N Engl J Med* 1998; **339**: 591–598.
3 Coustan-Smith E, Sancho J, Hancock ML, Boyett JM, Behm FG, Raimondi SC *et al*. Clinical importance of minimal residual disease in childhood acute lymphoblastic leukemia. *Blood* 2000; **96**: 2691–2596.
4 Borowitz MJ, Devidas M, Hunger SP, Bowman WP, Carroll AJ, Carroll WL *et al*. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factor. A Children's Oncology Group study. *Blood* 2008; **111**: 5477–5485.

© 2013 Macmillan Publishers Limited

5 Conter V, Bartram CR, Valsecchi MG, Schrauder A, Panzer-Grümayer R, Möricke A *et al.* Molecular response to treatment redefines all prognostic factors in children and adolescents with B-cell precursor acute lymphoblastic leukemia: results in 3184 patients of the AIEOP-BFM ALL 2000 study. *Blood* 2010; **115**: 3206–3214.
6 Schrappe M, Valsecchi MG, Bartram CR, Schrauder A, Panzer-Grümayer R, Möricke A *et al.* Late MRD response determines relapse risk overall and in subsets of childhood T-cell ALL: results of the AIEOP-BFM-ALL 2000 study. *Blood* 2011; **118**: 2077–2084.
7 Brüggemann M, Raff T, Flohr T, Gokbuget N, Nakao M, Droese J *et al.* Clinical significance of minimal residual disease quantification in adult patients with standard-risk acute lymphoblastic leukemia. *Blood* 2006; **107**: 1116–1123.
8 Patel B, Rai L, Buck G, Richards SM, Mortuza Y, Mitchell W *et al.* Minimal residual disease is a significant predictor of treatment failure in non T-lineage adult acute lymphoblastic leukaemia: final results of the international trial UKALL XII/ECOG2993. *Br J Haematol* 2010; **148**: 80–89.
9 van der Velden VHJ, Wijkhuijs JM, Jacobs DC, van Wering ER, van Dongen JJ. T cell receptor gamma gene rearrangements as targets for detection of minimal residual disease in acute lymphoblastic leukemia by real-time quantitative PCR analysis. *Leukemia* 2002; **16**: 1372–1380.
10 Verhagen OJ, Willemse MJ, Breunis WB, Wijkhuijs AJ, Jacobs DC, Joosten SA *et al.* Application of germline IGH probes in real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia. *Leukemia* 2000; **14**: 1426–1435.
11 van der Velden VHJ, Willemse MJ, van der Schoot CE, Hahlen K, van Wering ER, van Dongen JJ. Immunoglobulin kappa deleting element rearrangements in precursor-B acute lymphoblasticleukemia are stable targets fordetection of minimalresidualdiseaseby real-time quantitative PCR. *Leukemia* 2002; **16**: 928–936.
12 Bruggemann M, van der Velden VHJ, Raff T, Droese J, Ritgen M, Pott C *et al.* Rearranged T-cell receptor beta genes represent powerful targets for quantification of minimal residual disease in childhood and adult T-cell acute lymphoblastic leukemia. *Leukemia* 2004; **18**: 709–719.
13 van der Velden VHJ, Panzer-Grumayer ER, Cazzaniga G, Flohr T, Sutton R, Schrauder A *et al.* Optimization of PCR-based minimal residual disease diagnostics for childhood acute lymphoblastic leukemia in a multi-center setting. *Leukemia* 2007; **21**: 706–713.
14 Campana D, Coustan-Smith E. Advances in the immunological monitoring of childhood acute lymphoblastic leukaemia. *Best Pract Res Clin Haematol* 2002; **15**: 1–19.
15 Bjorklund E, Mazur J, Soderhall S, Porwit-MacDonald A. Flow cytometric follow-up of minimal residual disease in bone marrow gives prognostic information in children with acute lymphoblastic leukemia. *Leukemia* 2003; **17**: 138–148.
16 Dworzak MN, Fröschl G, Printz D, Mann G, Pötschger U, Mühlegger N *et al.* Prognostic significance and modalities of flow cytometric minimal residual disease in childhood acute lymphoblastic leukemia. *Blood* 2002; **99**: 1952–1958.
17 Pongers-Willemse MJ, Seriu T, Stolz F, d'Aniello E, Gameiro P, Pisa P *et al.* Primers and protocols for standardized detection of minimal residual disease in acute lymphoblastic leukemiausing immunoglobulin and T cell receptor gene rearrangements and TAL1 deletions as PCR targets: report of the BIOMED-1 CONCERTED ACTION: investigation of minimal residual disease in acute leukemia. *Leukemia* 1999; **13**: 110–118.
18 van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003; **17**: 2257–2317.
19 van der Velden VHJ, Cazzaniga G, Schrauder A, Hancock J, Bader P, Panzer-Grumayer ER *et al.* Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. *Leukemia* 2007; **21**: 604–611.
20 Brüggemann M, Schrauder A, Raff T, Pfeifer H, Dworzak M, Ottmann OG *et al.* Standardized MRD quantification in European ALL trials: proceedings of the Second International Symposium on MRD assessment in Kiel, Germany, 18-20 September 2008. *Leukemia* 2010; **24**: 521–535.
21 Robillard N, Cavé H, Méchinaud F, Guidal C, Garnache-Ottou F, Rorhlich PS *et al.* Four-color flow cytometry bypasses limitations of polymerase chain reaction for minimal residual disease detection in certain subsets of children with acute lymphoblastic leukemia. *Haematologica* 2005; **90**: 1516–1523.
22 Dworzak MN, Gaipa G, Ratei R, Veltroni M, Schumich A, Maglia O *et al.* Standardization of flow cytometric minimal residual disease evaluation in acute lymphoblastic leukemia: multicentric assessment is feasible. *Cytometry B Clin Cytom* 2008; **74**: 331–340.
23 Irving J, Jesson J, Virgo P, Case M, Minto L, Eyre L *et al.* Establishment and validation of a standard protocol for detection of minimal residual disease in B lineage childhood acute lymphoblastic leukemia by flow cytometry in a multicenter setting. *Haematologica* 2009; **94**: 870–874.
24 Björklund E, Matinlauri I, Tierens A, Axelsson S, Forestier E, Jacobsson S *et al.* Quality control of flow cytometry data analysis for evaluation of minimal residual disease in bone marrow from acute leukemia patients during treatment. *J Pediatr Hematol Oncol* 2009; **31**: 406–415.
25 Neale GA, Coustan-Smith E, Stow P, Pan Q, Chen X, Pui CH *et al.* Comparative analysis of flow cytometry and polymerase chain reaction for the detection of minimal residual disease in childhood acute lymphoblastic leukemia. *Leukemia* 2004; **18**: 934–938.
26 Malec M, van der Velden VHJ, Bjorklund E, Wijkhuijs JM, Soderhall S, Mazur J *et al.* Analysis of minimal residual disease in childhood acute lymphoblastic leukemia: comparison between RQ-PCR analysis of Ig/TcR gene rearrangements and multicolour flow cytometric immunophenotyping. *Leukemia* 2004; **18**: 1630–1636.
27 Kerst G, Kreyenberg H, Roth C, Well C, Dietz K, Coustan-Smith E *et al.* Concurrent detection of minimal residual disease (MRD) in childhood acute lymphoblastic leukaemia by flow cytometry and real-time PCR. *Br J Haematol* 2005; **128**: 774–782.
28 Bene MC, Castoldi G, Knapp W, Ludwig WD, Matutes E, Orfao A *et al.* Proposals for the immunological classification of acute leukemias. *Leukemia* 1995; **9**: 1783–1786.
29 Pui CH, Campana D, Pei D, Bowman WP, Sandlund JT, Kaste SC *et al.* Treating childhood acute lymphoblastic leukemia without cranial irradiation. *N Engl J Med* 2009; **360**: 2730–2741.
30 Coustan-Smith E, Song G, Clark C, Key L, Liu P, Mehrpooya M *et al.* New markers for minimal residual disease detection in acute lymphoblastic leukemia. *Blood* 2011; **117**: 6267–6276.
31 Thörn I, Forestier E, Botling J, Thuresson B, Wasslavik C, Björklund E *et al.* Minimal residual disease assessment in childhood acute lymphoblastic leukaemia: a Swedish multi-centre study comparing real-time polymerase chain reaction and multicolour flow cytometry. *Br J Haematol* 2011; **152**: 743–753.
32 Gaipa G, Cazzaniga G, Panzer-Grumayer RE, Veltroni M, Karawajew L, Silvestri D *et al.* Time point-dependent concordance of flow cytometry and RQ-PCR in the MRD detection in childhood ALL: the experience of the AIEOP-BFM- ALL MRD study group. *Blood* 2008; **112**: abstract 700.
33 Stow P, Key L, Chen X, Pan Q, Neale GA, Coustan-Smith E *et al.* Clinical significance of low levels of minimal residual disease at the end of remission induction therapy in childhood acute lymphoblastic leukemia. *Blood* 2010; **115**: 4657–4663.
34 Panzer-Grumayer ER, Schneider M, Panzer S, Fasching K, Gadner H. Rapid molecular response during early induction chemotherapy preditcs a good outcome in childhood acute lymphoblastic leukemia. *Blood* 2000; **95**: 790–794.
35 Coustan-Smith E, Sancho J, Behm FG, Hancock ML, Razzouk BI, Ribeiro RC *et al.* Prognostic importance of measuring early clearance of leukemic cells by flow cytometry in childhood acute lymphoblastic leukemia. *Blood* 2002; **100**: 52–58.
36 Basso G, Veltroni M, Valsecchi MG, Dworzak MN, Ratei R, Silvestri D *et al.* Risk of relapse of childhood acute lymphoblastic leukemia is predicted by flow cytometric measurement of residual disease on day 15 bone marrow. *J Clin Oncol* 2009; **27**: 5168–5174.
37 Szczepanski T, Willemse MJ, van Wering ER, van Weerden JF, Kamps WA, van Dongen JJ. Precursor-B-ALL with D(H)-J(H) gene rearrangements have an immature immunogenotype with a high frequency of oligoclonality and hyperdiploidy of chromosome 14. *Leukemia* 2001; **15**: 1415–1423.
38 Dworzak MN, Gaipa G, Schumich A, Maglia O, Ratei R, Veltroni M *et al.* Modulation of antigen expression in B-cell precursor acute lymphoblastic leukemia during induction therapy is partly transient: evidence for a drug-induced regulatory phenomenon. Results of the AIEOP-BFM-ALL-FLOW-MRD-Study Group. *Cytometry B Clin Cytom* 2010; **78**: 147–153.
39 Roshal M, Fromm JR, Winter S, Dunsmore K, Wood BL. Immaturity associated antigens are lost during induction for T cell lymphoblastic leukemia: implications for minimal residual disease detection. *Cytometry B Clin Cytom* 2010; **78**: 139–146.
40 Eckert C, Flohr T, Koehler R, Hagedorn N, Moericke A, Stanulla M *et al.* Very early/early relapses of acute lymphoblastic leukemia show unexpected changes of clonal markers and high heterogeneity in response to initial and relapse treatment. *Leukemia* 2011; **25**: 1305–1313.

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)

© 2013 Macmillan Publishers Limited